



AN EFFICIENT OPTIMIZATION BASED LUNG CANCER PRE-DIAGNOSIS SYSTEM WITH AID OF FEED FORWARD BACK PROPAGATION NEURAL NETWORK (FFBNN)

¹K.BALACHANDRAN, ²DR. R. ANITHA

¹Associate Professor, Department of Computer Science and Engineering, Christ University, Bengaluru

²Director, Department of Master of Computer Applications, K. S. Rangasamy College of Technology, Tiruchengode, Tamilnadu, India

E-mail: ¹balachandran0563@gmail.com

ABSTRACT

World Health Organization (WHO) reports that worldwide 7.6 million deaths are caused by cancer each year. Uncontrollable cell development in the tissues of the lung is called as lung cancer. These uncontrollable cells restrict the growth of healthy lung tissues. If not treated, this growth can spread beyond the lung in the nearby tissue called metastasis and, form tumors. In order to preserve the life of the people who are suffered by the lung cancer disease, it should be pre-diagonized. So there is a need of pre diagnosis system for lung cancer disease which should provide better results. The proposed lung cancer prediagnosis technique is the combination of FFBNN and ABC. By using the Artificial Bee Colony (ABC) algorithm, the dimensionality of the dataset is reduced in order to reduce the computation complexity. Then the risk factors and the symptoms from the dimensional reduced dataset are given to the FFBNN to accomplish the training process. In order to get higher accuracy in the prediagnosis process, the FFBNN parameters are optimized using ABC algorithm. In the testing process, more data are given to well trained FFBNN-ABC to validate whether the given testing data predict the lung disease perfectly or not.

Keywords: *Artificial Bee Colony (ABC) Algorithm, Feed Forward Back Propagation Neural Network (FFBNN), Risk factor and symptoms, Dimensionality reduction*

1. INTRODUCTION

Lung cancer is one of the [19] [1] commonest cancers in the industrialized world, and persons with this grave disease must deal not only with the physical effects but also with the psychosocial aspects [1]. Lung cancer [17] is a disease of abnormal cells multiplying and growing into a tumor [3]. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time [5]. Among different types of cancer the lung cancer is the most aggressive and best practice to its accurate prognosis is the determination of the current stage of the disease. Three main factors in cancer staging are primary tumor, regional lymph nodes and metastasis [4]. One of the most important and difficult tasks a doctor has to carry out is the detection and diagnosis of cancerous lung nodules from x-ray image's result [6].

Given that lung cancer is one of the common cancers world-wide, the implications of focusing on quality of life as well as survival require to be

understood [2]. Early detection is the most important for reducing the death due to lung cancer [20] [7]. The early detection of the lung cancer [14] [8] is a challenging problem [18], due to both the structure of the cancer cells and the stained methods which are used in the preparation of the sputum cells [8]. Now there are three main methods to use for diagnosis of lung cancer: biochemical diagnosis (serology and immunology), imaging diagnosis, and cytology histology diagnosis [9]. Previous researches have shown that the major types of lung cancer are associated with cigarette smoking [10]. A lot of research has been done on the detection of areas with high cancer incidence by domestic and foreign researchers, but the research did not relieve the increasing lung cancer incidence and mortality rates [11].

Most of the techniques are detecting the lung cancer in its advanced stages, where the patients' chance of survival is very low. Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages. Image processing and data mining techniques provide a good quality



tool for improving the manual analysis [13]. Medical Data Mining is a promising area of computational intelligence [16][15] applied to an automatically analyze patients records aiming at the discovery of new knowledge useful for medical decision-making [12]. Applying data mining techniques to cancer data is useful to rank and link cancer attributes to the survival outcome. Further, accurate outcome prediction can be extremely useful for doctors and patients to not only estimate survivability, but also aid in decision making to determine the best course of treatment for a patient, based on patient specific attributes, rather than relying on personal experiences, anecdotes, or population-wide risk assessments [15].

In this work, a lung cancer pre-diagnosis system is proposed to diagnose the lung cancer using FFNN-ABC technique. After gathering the lung cancer datasets, from the medical centers, the dimensionality of the dataset is reduced using ABC algorithm. After that the risk factors and symptoms causing lung cancer present in the dimensionality reduced lung cancer dataset are given to FFNN for training purpose. To get more accurateness in the diagnosis process, ABC is used in the training process. The rest of the paper is organized as follows: Section 2 reviews the related works with respect to the proposed method. Section 3 discusses about the proposed technique. Section 4 shows the experimental result of the proposed technique and section 5 concludes the paper.

2. RELATED WORKS

Fang *et al.* [21] have used a network-based biomarker detection approach coupled with gene set enrichment analysis to identify and validate genes associated with lung cancer and relating pathways. Results revealed a panel of new and unexpected genes with potential physiological functions in cigarette smoking or lung cancer using the coupled discovery strategy, in addition to those consistent with previous findings in cigarette smoking and lung cancer. Thus, it was indicated that disease-specific network biomarkers, interaction between genes/proteins, or cross-talking of pathways provide more specific values for the development of precision therapies for lung.

Berni *et al.* [22] have made a study to compare various delays in a rapid outpatient diagnostic program (RODP) for suspected lung cancer patients with those described in literature and with guideline recommendations, to investigate the effects of referral route and symptoms on delays, and to establish whether delays were

related to disease stage and outcome. Patient characteristics, tumor stage and different delays were analyzed. Medical charts of 565 patients were retrieved. 290 patients (51.3%) were diagnosed with lung cancer, 48 (8.5%) with another type of malignancy, and in 111 patients (19.6%) the radiological anomaly was diagnosed as non-malignant. Patients presenting with hemoptysis had shorter first line delays. An RODP was resulted in a timely diagnosis well within guideline recommendations. Patient and first line delay account for most of total patient delay. Within the limitations of this retrospective study, they found no association with disease stage or survival.

Nancy *et al.* [23] have presented an implication network-based approach to the identification of a smoking-associated 6-gene signature that was co-expressed with major Non-Small Cell Lung Cancer (NSCLC) signaling pathways. The methodology contained the following steps: (1) identifying genes significantly associated with lung cancer survival; (2) selecting candidate genes which are differentially expressed in smokers versus non-smokers from the survival genes identified in Step 1; (3) from these candidate genes, constructing gene co-expression networks based on prediction logic for the smoker group and the non-smoker group, respectively; (4) identifying smoking-mediated differential components, i.e., the unique gene co-expression patterns specific to each group; and (5) from the differential components, identifying genes directly co-expressed with major lung cancer signaling hallmarks. The pathway-based approach identified a smoking-associated 6-gene signature that predicts lung cancer risk and survival. This gene signature had potential clinical implications in the diagnosis and prognosis of lung cancer in smokers.

Tran [24] have presented a method that can find cost-effective biological markers for a perfect predictive accuracy of Non-Small-Cell Lung Cancer (NSCLC) lung cancers. As cancers are complicated, one can only predict the status using a combination of many genes. They have discovered a small set of nine gene-signatures from the dataset of 12,600 gene expression profiles of NSCLC which act like an inference basis for NSCLC lung carcinoma and used as genetic markers. These very small and previously unknown sets of biological markers gave an almost perfect predictive accuracy (99.75 %) for the diagnosis of the disease the sub-type of cancer. They presented a method that found genetic markers for sub-classification of NSCLC.

Generalized Loren z curve s and Gini ratios were used to overcome many challenges arose from datasets of gene -expression profiles. This method discovered genetic changes that occur in lung tumors using gene-expression profiles.

Polat *et al.* [25] have detected lung cancer using principles component analysis (PCA), fuzzy weighting preprocessing and artificial immune recognition system (AIRS). The system has three stages. First, dimensionality of lung cancer dataset that has 57 features was reduced to four features using principles component analysis. Second, a weighting scheme based on fuzzy weighting pre-processing was utilized as a pre-processing step before the main classifier. Third, artificial immune recognition system was used classifier. Experiments were conducted on the lung cancer dataset to diagnose lung cancer in a fully automatic manner. The obtained classification accuracy of system was 100% and it was very promising with regard to the other classification applications.

3. PROPOSED METHODOLOGY

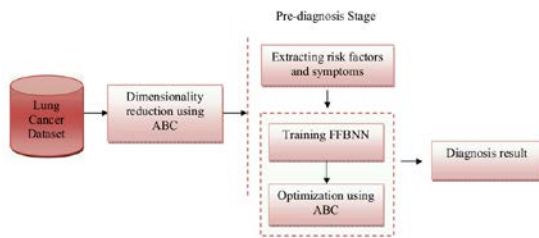


Fig. 1. Architecture Of The Proposed Lung Cancer Pre-Diagnosis System

The architecture of proposed lung cancer prediagnosis system is shown in Fig1. In the beginning, the dimensionality of the given dataset is reduced using Artificial Bee Colony (ABC) algorithm. After the dimensionality reduction, the reduced dataset are given as the input to the prediagnosis stage. Here, the high risk factors and symptoms which cause lung cancer are extracted from the dimensionality reduced dataset. These features are given as the input to the Feed Forward Back Propagation Neural Network (FFBNN). While training, the FFBNN parameters are optimized using ABC algorithm. During the testing process, more number of patient’s data is given to well trained FFBNN-ABC to validate whether the given testing data predict the lung disease perfectly or not.

The proposed lung cancer pre-diagnosis system consists of the following phases:

- ❖ Dimensionality reduction
- ❖ Pre-diagnosis

3.1 Dimensionality Reduction

At first the lung cancer dataset (D) is obtained from the medical center which is the size of $(R \times C)$. $R = 1, 2, \dots, n$ denotes the number of rows of the dataset and $C = 1, 2, \dots, m$ denotes the number of columns in the dataset. The lung cancer dataset (D) contains the information about the patients those who have lung cancer, other cancer diseases and no cancer disease and also the corresponding risk factors and symptoms of those patients. By using the ABC algorithm the size of the dataset is reduced in order to decrease the computational complexity. The working procedure of ABC algorithm is shown in Fig2.

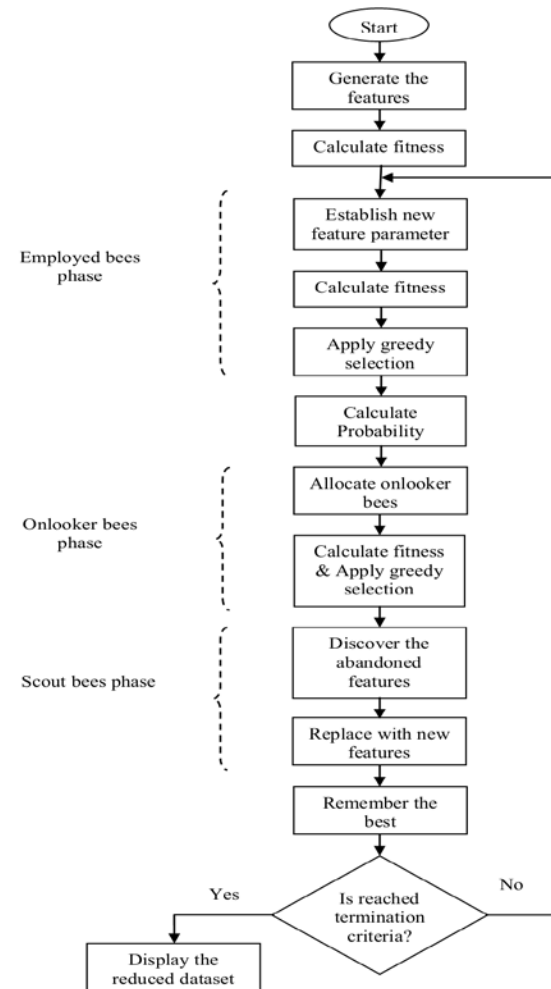


Fig.2. Flowchart For ABC Algorithm



ABC algorithm is a swarm based meta-heuristic algorithm which was enthused by the sharp foraging behavior of the honey bees. It consists of three components namely, employed bees, onlooker bees and scout bees. The employed bees are coupled with the food sources in the region of the hive and they transfer the data to the onlookers about the nectar quality of the food sources they are exploiting. Onlooker bees are looking the dance of the employed bees inside the hive to pick one food source to exploit according to the data provided by the employed bees. The employed bees whose food source is abandoned become Scout and seeking new food source arbitrarily. The number of food sources denotes the location of probable solutions of optimization problem and the nectar amount of a food source denotes the quality of the solution.

Initial Phase

First the population of the food sources x_i ($i=1,2,\dots,N$) are generated arbitrarily. N denotes the size of the population. This food sources contains the row (R_i) values of the dataset. This generation process is called as initialization process. To evaluate the best food source, the fitness value of the generated food sources is calculated using equation. (1)

$$Fitness\ function\ F(j) = \max(f_i) \quad (1)$$

In Eq. (1) $F(j)$ is the fitness function of j^{th} parameter.

$$f_i = \frac{n_c}{n} \times 100 \quad (2)$$

Where, n_c - Correct classification, n - Total number of classification. After the calculation of fitness value, the iteration is set to 1. After that, the phase of employed bee is carried out.

Employed Bee Phase

In the employed bee phase, new population parameters are generated using the below equation,

$$V_{i,j} = x_{i,j} + \phi_{ij}(x_{i,j} - x_{k,j}) \quad (3)$$

Where, k and j is a random selected index, ϕ is randomly produced number in the range $[-1, 1]$ and $V_{i,j}$ is the new value of the j^{th} position. Then the fitness value is computed for every new generated population parameters of food sources. From the computed fitness value of the population, best population parameter is selected i.e. the

population parameter, which has the highest fitness value by applying greedy selection process. After selecting the best population parameter, probability of the selected parameter is computed using the equation (4).

$$P_j = \frac{F_j}{\sum_{j=1}^d F_j} \quad (4)$$

Where, P_j is the probability of the j^{th} parameter.

Onlooker Bee Phase

After computing the probability of the selected parameter, number of onlooker bees is estimated. Following, generate new solutions ($V_{i,j}$) for the onlooker bees from the solutions ($x_{i,j}$) based on the probability value (P_j). Then the fitness function is calculated for the new solution. Subsequently apply the greedy selection process in order to select the best parameter.

Scout Bee Phase

Determine the Abandoned parameters for the scout bees. If any abandoned parameter is present, then replace that with the new parameters discovered by scouts using the equation (4) and evaluate the fitness value. Then memorize the best parameters achieved so far. Then the iteration is incremented and the process is continued until the stopping criterion is reached. Finally, the reduced dataset D' ($D' = R' \times C$) is discovered.

3.2 Prediagnosis of Lung Cancer

To diagnosis the lung cancer, Feed Forward neural Network (FFBNN) is utilized. In the training phase, the risk factors and symptoms utilized from the reduced dataset are given as the input to the FFBNN. The neural network is well trained using these extracted risk factors and symptoms from the reduced lung cancer dataset in order to diagnosis the lung cancer. The neural network consists of n number of input units, h hidden units and one output unit. The structure of the FFBNN is given as below:

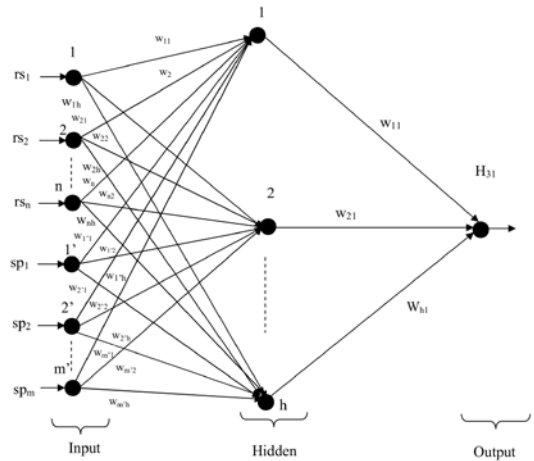


Fig.3. Diagram Of The FFBNN

1. For all the neurons, allot weights arbitrarily except for input neurons.
2. The bias function and activation function for the neural network is described below.

$$x(t) = \beta + \sum_{n=1}^H \left(w_m rs_{t1} + w_m rs_{t2} + \dots + w_m rs_{tm} \right) \quad (5)$$

$$x(a) = \frac{1}{1 + e^{-x(t)}} \quad (6)$$

In bias function $rs_{t1}, rs_{t2} \dots rs_{tm}$ are the extracted risk factors and $sp_{t1}, sp_{t2} \dots sp_{tm}$ are the extracted symptoms from the reduced dataset. The activation function for the output layer is given in Eq. (27).

3. Get the learning error.

$$Er = \frac{1}{h} \sum_{n=0}^{h-1} de_n - ac_n \quad (7)$$

Er is the FFBNN network output, de_n and ac_n are the desired and actual outputs and h is the total number of neurons in the hidden layer.

3.2.1.1 Error Minimization

Weights are allocated to the hidden layer and output layer neurons by randomly chosen weights. The input layer neurons have a constant weight.

1. Determine the bias function and the activation function.
2. Calculate BP error for each node and update the weights as follows:

$$w(tn) = w(tn) + \Delta w(tn) \quad (8)$$

$\Delta w(tn)$ is obtained as,

$$\Delta w(tn) = \delta \cdot x(tn) \cdot Be \quad (9)$$

Where δ is the learning rate, which normally ranges from 0.2 to 0.5, and Be is the Back Propagation error.

3. Then repeat the steps (2) and (3) until the BP error gets minimized. The process is repeated until it satisfies $Be < 0.1$.

4. The error gets minimized to a minimum value the FFBNN is well trained for performing the testing phase.

Then the result of the neural network (y) is compared with the threshold value (τ). If it satisfies the threshold value it denotes presence of lung cancer.

$$result = \begin{cases} lung\ cancer, & y \geq \tau \\ not\ lung\ cancer, & y < \tau \end{cases}$$

In order to get higher accuracy and effective performance in diagnosis of the lung cancer, the FFBNN parameters (w_m, β) are optimized using ABC.

3.2.2 Optimization of FFBNN parameters by ABC

A detailed description about the ABC algorithm and the working procedure is given in section (3.1) already. Here we are using the ABC algorithm again for optimizing the parameters of FFBNN while training itself to get efficient diagnosis result. The working procedure is same as given in section (3.1) excepting fitness function. The FFBNN parameters (w_m, β) are optimized using ABC. The fitness function preferred here is eqn. (2). This optimization of FFBNN parameters using ABC gives higher diagnosis result and effective performance. In the testing process, various risk factors and symptoms are given to well trained FFBNN-ABC to check whether the testing data will predict the lung disease exactly or not.

4. EXPERIMENTAL RESULTS

The proposed lung cancer pre-diagnosis system with FFBNN-ABC is implemented in the working platform of MATLAB (version 7.12) with machine configuration as follows

Processor: Intel core i7
OS: 3.20 GHz
CPU speed: Windows 7
RAM: 4GB

The proposed lung cancer prediagnosis technique is the combination of FFBNN and ABC. In order to



reduce the computation complexity and get higher performance, the dimensionality of the dataset (D) is reduced with the help of the well known optimization algorithm ABC. Then the risk factors and the symptoms from the dimensional reduced dataset (D') are given to the FFBNN to achieve the training process. So as to get more accuracy in the process of prediagnosis, the FFBNN parameters are optimized using ABC algorithm. In the testing process, more data are given to the well trained FFBNN-ABC to validate the performance of the proposed technique. The performance of the proposed lung cancer pre-diagnosis technique is evaluated using the lung cancer dataset which contains risk factors and symptoms and the proposed technique's performance is compared with the FFBNN- GA (Genetic Algorithm), FFBNN-PSO (Particle Swarm Optimization) and FFBNN.

4.1 Performance Analysis

The performance of our proposed lung cancer pre-diagnosis system is analyzed by using the statistical measures which is given in [26]. To carry out the performance examination process we make use of lung cancer dataset which has the patients list those who have lung cancer, other cancer and no cancer and also the corresponding risk factors and symptoms of those patients. The size of the given dataset is 44 x77. By using the familiar ABC algorithm, the given lung cancer dataset size is reduced. The risk factors and the symptoms obtained from the reduced dataset are utilized to validate the performance of the proposed lung cancer pre-diagnosis technique and the performance of the proposed technique is compared with the other optimization techniques such as FFBNN-GA, FFBNN-PSO and FFBNN. In addition, the statistical measures of our proposed and existing systems are given in Table 1 and 2.

Table. 1. Statistical Measures

Technique	TP	TN	FP	FN
Proposed FFBNN-ABC	15	3	0	2
FFBNN-GA	10	2	5	3
FFBNN - PSO	14	2	1	3
FFBNN	8	2	7	3

Statistical measures such as TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative) are given in the above table1 for proposed FFBNN – ABC technique and the other

optimization techniques such as, FFBNN-GA, FFBNN-PSO and FFBNN techniques.

True Positive- correctly identified as the lung cancer

True Negative-correctly identified as other disease

False Positive-incorrectly identified as the lung cancer

False Negative-incorrectly identified as the other disease

By using the TP, TN, FP and FN values the other statistical measures like accuracy, sensitivity and specificity values are calculated and it is given in table2.

Table. 2.Performance Of Our Proposed Lung Cancer Pre-Diagnosis Technique And Other Optimization Technique Such As FFBNN-GA, FFBNN-PSO And FFBNN

Measures	Proposed FFBNN-ABC	FFBNN-GA	FFBNN-PSO	FFBNN
Accuracy	90	60	80	50
Sensitivity	88	77	82	0
Specificity	100	29	67	22
FPR	0.0	71.4	33.3	77.8
PPV	100	67	93	53
NPV	60	40	40	40
FDR	0	33	7	47
MCC	66.4	6.8	39.1	0.0

In table2, accuracy, sensitivity and specificity measures are given. The accuracy of the proposed technique is 90%. FFBNN-GA, FFBNN-PSO, FFBNN has 60%, 80% and 50% of accuracy respectively. When compared to the proposed FFBNN-ABC technique, FFBNN-PSO has 10% lesser accuracy; FFBNN-GA has 30% lesser accuracy and FFBNN has 40% lower accuracy. It represents the proposed technique predict the lung cancer more accurately than the other techniques. While seeing the sensitivity measure, the proposed technique has 88% sensitivity. And FFBNN-GA has 77% of sensitivity; FFBNN-PSO has 82% of sensitivity; FFBNN has 0% of sensitivity. High percentage of sensitivity indicates the good performance of the technique. Thus by seeing the sensitivity measures, the proposed technique has higher performance since the proposed technique's sensitivity is 11% higher than FFBNN-GA; 6% higher than FFBNN-PSO and 88% higher than FFBNN. Specificity is another one measurement to validate the performance of the proposed technique. FFBNN- GA has 29% of accuracy; FFBNN-PSO has 67 % of specificity. FFBNN has 22% of specificity. But the proposed technique has 100%

of specificity. It is about 33 – 78% higher than the other techniques such as FFBNN-GA, FFBNN-PSO and FFBNN. It indicates that the proposed technique has higher performance than the other techniques.

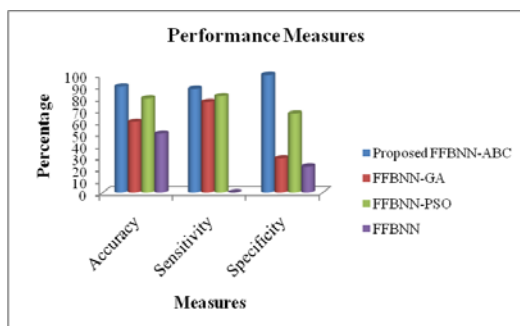


Fig. 4. Proposed, FFBNN- GA, FFBNN-PSO And FFBNN Techniques Performance Outcomes In Terms Of Accuracy, Sensitivity And Specificity

In Fig 4, the accuracy, sensitivity and specificity of the proposed technique is compared with the other techniques. By seeing the graph, the accuracy of the proposed technique is significantly higher than the FFBNN-GA, FFBNN-PSO, and FFBNN. Similarly the sensitivity and specificity measures of the proposed FFBNN-ABC technique are also notably higher than the other optimization techniques. It indicates the performance and the exactness of the proposed technique is higher than the other techniques. Our proposed FFBNN-ABC lung cancer pre-diagnosis system has achieved 90% of accuracy, 88% of sensitivity and 100% of specificity respectively. It indicates that our proposed lung cancer pre-diagnosis system accomplished better performance when compared to the other lung cancer diagnosis techniques such as FFBNN- GA, FFBNN-PSO and FFBNN. The comparison result shows that the proposed FFBNN-ABC lung cancer pre-diagnosis system has more precisely pre-diagnose lung cancer than the other techniques.

5. CONCLUSION

In this paper, we have proposed a lung cancer pre-diagnosis system with the aid of FFBNN and ABC. The proposed system was implemented and a huge set of test data's were utilized to analyze the outcomes of the proposed lung cancer pre-diagnosis system. Thus the proposed lung cancer pre-diagnosis system offers a significant tempo of accuracy, sensitivity and specificity. We can say that proposed method more precisely diagnosis the lung cancer from the given test data by seeing the

elevated rate of measurements. The comparison result shows that our proposed lung cancer pre-diagnosis system based on FFBNN-ABC has given high exactness than FFBNN- GA, FFBNN-PSO and FFBNN methods. Therefore by utilizing the FFBNN and ABC techniques, our proposed lung cancer pre-diagnosis system proficiently diagnoses the lung cancer.

REFERENCES

- [1] Kathrine Carlsen, Anders Bonde Jensen, Erik Jacobsen, Mark Krasnik and Christoffer Johansen, "Psychosocial aspects of lung cancer Review Article", *Lung Cancer*, Vol. 47, No. 3, pp. 293-300, 2005.
- [2] Ali Montazeri, Robert Milroy, David Hole, James McEwen and Charles R. Gillis, "Quality of life in lung cancer patients: As an important prognostic factor Original Research Article", *Lung Cancer*, Vol. 31, No. 2-3, pp. 233-240, 2001.
- [3] Mokhled S. Al-Tarawneh, "Lung Cancer Detection Using Image Processing Techniques", *Leonardo Electronic Journal of Practices and Technologies*, Vol. 11, No. 20, pp. 147-158, 2012.
- [4] Almasslawi and Kabir , "Using non-rigid image registration and Thin-Plate Spline warping for lung cancer progression assessment", *In IEEE proceedings of International Conference on Computer Science and Automation Engineering, Shanghai*, pp. 506-510, 2011.
- [5] Chaudhary and Singh , "Lung Cancer Detection on CT Images by Using Image Processing", *In IEEE Proceedings of International Conference on Computing Sciences, Phagwara*, pp.142-146, 2012.
- [6] Abdullah and Mohamaddiah, "Development of cellular neural network algorithm for detecting lung cancer symptoms", *In IEEE Proceedings of International Conference on Biomedical Engineering and Sciences, Kuala Lumpur*, pp.138-143, 2010.
- [7] Wen-Lin Chang , Chia-Chen Chang and Wei-Ming Chen, "Computer-Aided Diagnosis System for Variance Estimation of 3D Ultrasonography Based on Gabor Filter", *In IEEE Proceedings of International Conference on Biomedical Engineering and Computer Science, Wuhan*, pp. 1-4, 2010.
- [8] Fatma Taher, Naoufel Werghi and Hussain Al-Ahmad, "A Thresholding Approach For Detection of Sputum Cell For Lung Cancer



- Early Diagnosis”, In *IEEE Proceedings of IET Conference on Image Processing, London*, pp. 1-6, 2012.
- [9] Yongjun WU, Na Wang, Hongsheng ZHANG, Lijuan Qin, Zhen YAN And Yiming WU, “Application of Artificial Neural Networks in the Diagnosis of Lung Cancer by Computed Tomography”, In *IEEE Proceedings of the Sixth International Conference on Natural Computation, China*, pp. 147-153, 2010.
- [10] Yi-Bin Wang, Yong-Mei Cheng, Shao-Wu Zhang and Wei Chen, “A Seed-based Approach to Identify Risk Disease Sub-networks in Human Lung Cancer”, In *IEEE Proceedings of the 6th International Conference on Systems Biology, Xi'an*, pp. 135-141, 2012.
- [11] Cuiping Wang, Xinhui Li, Long Dai, Guoqin Zhang, Liling Gao and Qianjun Zhao, “Analysis of Lung Cancer Gender Differences and Age Structure in the High Prevalence Areas of Xiamen”, In *IEEE Proceedings of International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou*, pp. 154 - 159, 2011.
- [12] Ada and Rajneet Kaur, “A Study of Detection of Lung Cancer Using Data Mining Classification Techniques”, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, No.3, pp. 131-134, 2013.
- [13] Ada and Rajneet Kaur, “Using Some Data Mining Techniques to Predict the Survival Year of Lung Cancer Patient”, *International Journal of Computer Science and Mobile Computing*, Vol. 2, No. 4, pp.1 – 6, 2013.
- [14] Ayman El-Baz, Georgy Gimelfarb, Robert Falk and Mohamed Abo El-Ghar, “A New Cad System For Early Diagnosis Of Detected Lung Nodules”, In *IEEE Proceedings of International Conference on Image Processing, San Antonio, TX*, pp. II-461 - II-464, 2007.
- [15] Ankit Agrawal, Sanchit Misra, Ramanathan Narayanan, Lalith Polepeddi and Alok Choudhary, “A Lung Cancer Outcome Calculator Using Ensemble Data Mining on SEER Data”, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, United States*, pp. 51-59, 2011.
- [16] Zakaria Suliman Zubi and Rema Asheibani Saad, “Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer”, *Proceedings of the 10th WSEAS international conference on Artificial intelligence Knowledge Engineering and Data Bases*, pp. 32-37, 2011.
- [17] Ying-Wooi Wan and Guo, “Constructing gene-expression based survival prediction model for Non-Small Cell Lung Cancer (NSCLC) in all stages and early stages”, In *IEEE Proceedings of International Conference on Bioinformatics and Biomedicine Workshop, Washington*, pp. 338, 2009.
- [18] Taher, Fatma and Sammouda, “Lung cancer detection by using artificial neural network and fuzzy clustering methods”, In *IEEE Proceedings of GCC Conference and Exhibition, Dubai*, pp. 295-298, 2011.
- [19] Liu Lu and Liu Wanyu, "A Method of Pulmonary Nodules Detection with Support Vector Machines", In *IEEE Proceedings of the Eighth International Conference on Intelligent Systems Design and Applications, Kaohsiung*, pp. 32-35, 2008.
- [20] Qian Yu, Lifeng He, Nakamura, Suzuki, K and Yuyan Chao, “A multilayered partitioning image registration method for chest-radiograph temporal subtraction”, In *IEEE Proceedings of International Conference on Computer Science and Information Processing, Xi'an, Shaanxi*, pp. 181-184, 2012.
- [21] Xiacong Fang, Michael Netzer, Christian Baumgartner, Chunxue Bai and Xiangdong Wang, “Genetic network and gene set enrichment analysis to identify biomarkers related to cigarette smoking and lung cancer”, *Cancer Treatment Reviews*, Vol. 39, No. 1, pp. 77-88, 2013.
- [22] Pepijn Brocken, Berni Kiers, Monika G. Looijen-Salamon, Richard Dekhuijzen, Chantal Smits-vander Graaf, Liesbeth Peters-Bax, Lioe-Fee de Geus-Oei and Henricus F.M. van der Heijden, “Timeliness of lung cancer diagnosis and treatment in a rapid outpatient diagnostic program with combined 18FDG-PET and contrast enhanced CT Scanning”, *Lung Cancer*, Vol. 75, No. 3, pp. 36- 341, 2012.
- [23] Nancy Lan Guo and Ying-Wooi, “Wan Pathway-based identification of a smoking associated 6-gene signature predictive of lung cancer risk and survival”, *Artificial Intelligence in Medicine*, Vol. 55, No. 2, pp. 97-105, 2012.
- [24] Quoc-Nam Tran, “A novel method for finding non-small cell lung cancer diagnosis biomarkers”, *International Conference on Bioinformatics and Computational Biology, Las Vegas, NV, USA*, pp. 18-21, 2011.



- [25] Kemal Polat and Salih Gunes, “Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer”, *Expert Systems with Applications*, Vol.34, No. 1, pp. 214–221, 2008.
- [26] https://en.wikipedia.org/wiki/Sensitivity_and_specificity.